Hierarchical Network Exploration using Gaussian Mixture Models

James Mathews, Saad Nadeem, Maryam Pouryahya,
Zehor Belkhatir, Joseph O. Deasy, Allen Tannenbaum*

ABSTRACT

We present a framework based on optimal mass transport to construct for a given network a supernode hierarchy which can be used for unsupervised interactive data exploration. Specifically, given a network and a set of numerical data samples for each node, we calculate a new computationally-efficient comparison metric between Gaussian Mixture Models, Gaussian Mixture Transport distance, to determine a series of node-merging simplifications of the network. The method has its basis in the local connection structure of the network, as well as the joint distribution of the data associated with neighboring nodes. The computation of GMT for all adjacent edge pairs in the network is highly parallelizable and performed only once prior to exploration, allowing real-time interactivity.

1. INTRODUCTION

The growing importance of complex networks has been documented in a huge and growing literature, now being referred to as the field of *network science* [1]. A key problem is the representation of network data in a readily accessible format. Ideally the representation should be amenable to human-in-the-loop, interactive, exploratory data analysis.

Supernode methods have been used previously to compress very large networks to a desired level of resolution, mainly toward the goal of improving the computational performance of community detection algorithms [2; 3]. The key idea of supernodes is to group nodes into modules and consider the new network comprised of the connections between groups implied by the individual node connections. In its most naive form, the groups may be obtained by devising an edge weighting intended to measure similarity between neighboring nodes, and successively "collapsing" edges beginning with the greatest similarity to create a hierarchical representation.

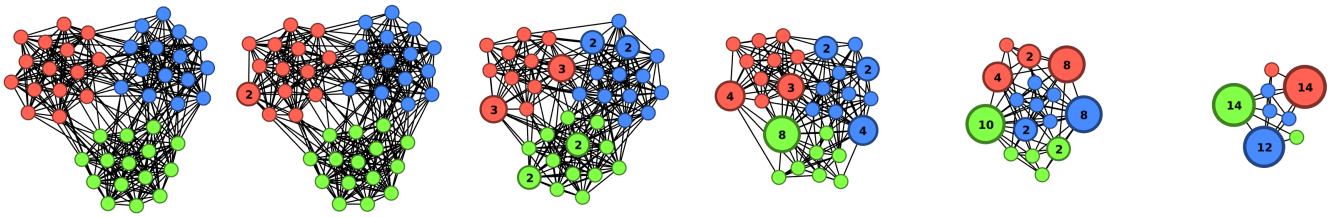
A natural improvement is to find approximations of a given network "from below" with gradually expanding small node subsets, an approach developed by Stanley *et al.* [2]. The authors randomly distribute seed nodes which are then expanded into supernodes using direct neighborhoods. In a more global approach, Yang *et al.* [3] present a method of supernode network representation involving explicit consideration of known prior constraints on the set of network topologies of interest for a low-complexity approximation to the given network satisfying the constraints. For a more detailed survey on this topic, refer to Besta *et al.* [4].

Unlike the method of [2] and in common with the method of [3], our proposed approach makes essential use of additional data beyond the network topology. However, rather than qualitative constraints as in [3], we assume numerical data along the nodes of the network. Data of this type is frequently available in practice when each node represents a variable of interest across a sample set, for example, mRNA expression of a gene (node) across a patient tissue sample population. On the other hand such data can also be generated from the underlying network topology if the topology is the primary variable of interest and sample analysis is not needed. The node weightings are interpreted as defining samples from the joint distribution of random variables associated with the nodes.

A very natural model for distributions is the Gaussian mixture used in many data processing and analysis applications [5]. In general, a ***mixture model*** is a weighted linear combination of distributions where each component represents a subpopulation. In particular, the Gaussian mixture model (GMM) is a weighted average of Gaussians. GMMs are popular given their versatility and overall simplicity in data representation. They are ubiquitous in statistics, hypothesis testing, decision theory, and machine learning. The idea is that real-world data may not be densely distributed on a high dimensional space, and instead is concentrated in a low dimensional subspace. Further, in many cases of interest, the data is sparsely distributed into a number of subgroups, and so differences within a given subgroup are not as important as those among the subgroups. Mixture models capture these properties, and this motivated the work of Chen *et al.* [6] modifying optimal mass transport (OMT) theory [7; 8] into a form suitable for Gaussian mixture models.

OMT provides the Gaussian mixture framework with a natural comparison metric between mixtures, and reciprocally mixtures provide a natural model with which to make the computation of OMT tractable. We call the resulting metric the ***Gaussian Mixture Transport*** (GMT) distance. Specifically, we use GMMs to quantify the functional role played by a node with respect to its neighbors in the network, and GMT distance is used to compare nodes based

*A. Tannenbaum is with the Departments of Computer Science and Applied Mathematics & Statistics, Stony Brook University, NY; email: allen.tannenbaum@stonybrook.edu



(a) (b) (c) (d) (e) (f)

Figure 1: A synthetic network with K=3 communities containing N=45 nodes total. The network was randomly generated to have within-community edge connectivity 0.08 out of a possible maximum of 0.68, and between-community edge connectivity 0.28 out of a possible maximum of 0.32. A random node weighting can be generated from the network by iterated neighborhood-averaging applied to an initial node weighting equal to 1 on a randomly selected node and 0 on every other node. 200 node weightings were generated in this manner. GMT-distance-based node similarity scores, calculated from the network and node weightings, determine a list of node-merge operations. The result is a hierarchical series of 44 simplified supernode networks. (a) The original network. (b)-(f) 5 selected supernode hierarhical levels from the series. Each supernode is labelled by the number of ordinary nodes it contains. *The supernode visualization is implemented as a user-friendly plugin in the Gephi software. The level of network simplification (supernode hierarchical level) is selected with a simple slider and shown in real-time.*

on their functional roles. Rather than constructing the supernode hierarchy by collapsing edges, we construct it using the GMT similarity metric. This provides a simplified version of the network for each given level of complexity. The simplified or compressed network represents a "projection" of the prior network which is most relevant according to the evidence observed in the data.

The contributions of this work are as follows:

1. We introduce a novel supernode hierarchy construction method that represents a network at a chosen level of complexity that merges edges by OMT-based statistical criteria on pairs of node-to-node relationships rather than criteria on pairs of nodes.
2. We demonstrate the effectiveness of the method in real-world scenarios by using the supernode networks to identify structural features in an mRNA expression lung cancer cell-line dataset.
3. We show that our parallelized implementation of the algorithm coupled with a visualization tool allows real-time interactivity compared with hierarchical methods requiring intensive recursive calculation for every layer.

This paper is organized as follows. Section 2 describes the steps of the computation in general, explaining the details in terms of a synthetic network example illustrated in the various figures. More precise pseudocode is provided in the algorithms. The key theoretical background from [6] is recalled in Section 3. Section 4 shows the results of a comparison against the method of Stanley *et al.* [2] and analysis of the supernode hierarchies calculated on the Kyoto Encyclopedia of Genes and Genomes (KEGG) cancer gene network using lung cancer cell-line mRNA expression data made available as part of the Genomics of Drug Sensitivity in Cancer (GDSC) project [9]. Finally, Section 5 describes further aims of our work.

2. METHODS

To illustrate the construction of GMT distance-based supernode hierarchies, we show the steps of the construction applied to a concrete example, a network synthesized in order to have a known community structure. The edge density is high within the communities, and low between communities. Node weightings are randomly generated in terms of the network topology by neighbor averaging, with one example of such a weighting depicted in Figure 2 (c). A preview of the series of network simplifications is shown in Figure 1.

2.1 Gaussian Mixture Transport distance

A naive approach would group nodes together based on similar properties, for example, by making comparisons between the univariate distributions of the weight data associated with each node. Comparison in this case is a classical topic, addressed, for instance, by the Kolmogorov-Smirnov test.

For a higher-order approach, we consider the bivariate distributions associated with each edge, meaning the joint distributions of the variables associated with the two endpoint nodes, then compare the two bivariate distributions for pairs of adjacent edges A-C and B-C (in this case, adjacent along node C). If the distributions are similar, A and B will be considered for membership in the same supernode for the specific reason that there are connections in the network which provide evidence for similarity of the functional role of A and the functional role of B.

How is the similarity between bivariate distributions quantified? The Bhattacharyya distance [10] is one similarity measure between distributions in higher dimensions, but a direct calculation of this distance tends to require rasterization of the space involved and may be prohibitively costly to compute. The theory developed by Chen *et al.* [6] for Gaussian Mixture Models provides a nearly closed-form alternative which is well-suited to the task. First, the distributions are approximated by Gaussian mixtures with a given number of subpopulations. The mixture weights are interpreted as probability distributions on the discrete set of subpopulations, which are themselves compared using the

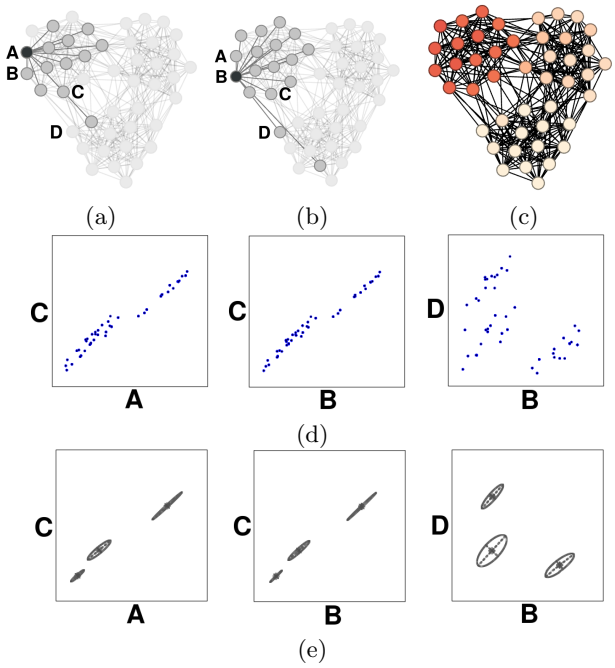


Figure 2: Illustration of the calculation of the GMT similarity for selected nodes in the network depicted in Figure 1. Nodes labelled A and B were selected from the same community. (a) The neighborhood of node A. (b) The neighborhood of node B. Both of these neighborhoods contain node C, while node D is only in the neighborhood of B. (c) One of the 200 randomly generated node weightings described in Figure 1. (d) The joint distributions of the weight values for node pairs (C,A), (C,B), and (D,B). (e) The 3-population 2D Gaussian Mixture Models for these joint distributions. The GMT distance between model (C,A) and model (C,B) is 3.3×10^{-7}. This value is typical of the GMT distance among models (C′, A) and (C′, B) where C′ is any other node besides C intervening between A and B, so the GMT similarity between A and B, which averages these values, will be small. This reflects the fact that A and B belong to a highly connected community. By contrast the GMT distance between models (C,B) and (C,D) is 1.0×10^{-4}, a value typical among C' intervening between B and D. So the GMT similarity between B and D will be relatively large, reflecting the fact that B and D were selected from different highly connected communities. In this example it is not essential that C was selected from the same community as A and B; though the typical mutual neighbor of A and B will be contained in the same community, like C, the same conclusions follow when C is selected from elsewhere in the network.

optimal mass transport metric or Earth Mover's Distance (EMD). For this calculation of the EMD, the cost function corresponding to motion from the discrete point labelling a subpopulation of the first mixture to a discrete point labelling a subpopulation of the second mixture is the actual optimal mass transport distance between the corresponding ordinary Gaussian distributions. For a detailed description, see Algorithm 2 summarizing the formulae of [6] and the OMT Background section 3. An example of the GMT distance comparison of node weight distributions is illustrated in Figure 2.

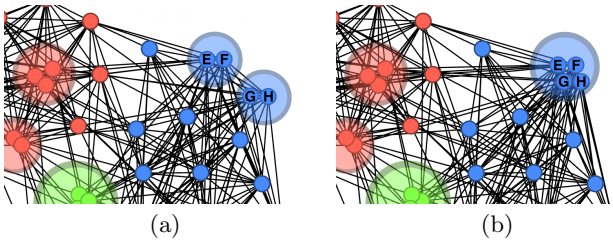


Figure 3: A supernode merge operation in the simplification illustrated in Figure 1. (a) Nodes E and F have already merged into a supernode, and nodes G and H have also already merged into a supernode. (b) The two supernodes merge into one supernode on the basis of GMT similarity scores already calculated between pairs (E,G), (E,H), (F,G) and (F,H). This particular merge takes place at this step because one of these 4 pairs has the smallest GMT similarity value among unmerged remaining node pairs.

2.2 Supernode hierarchy

The network simplification procedure is basically to merge GMT-similar *edges* which are adjacent along some node, rather than merging nodes directly. Note that for three random variables A, B, and C, if the joint distribution of A and C is very close to the joint distribution of B and C, it follows that the individual distributions of A and of B are very close. This means that if two edges A-C and B-C are to be merged because of small GMT distance, the merging of nodes A and B is not contradicted by evidence from these individual distributions. On the other hand, other, unrelated edge pairs may be inadvertently collapsed as a result. To guard against this possibility, we instead merge nodes A and B on the basis of the *average* of all GMT distances of edge pairs based on endpoints A and B which are adjacent along some third node. That is, we average over the intersection of the neighborhood of A and the neighborhood of B. This average is termed the GMT similarity between A and B.

After all GMT similarities are calculated according to Algorithm 1, the merge operations begin, starting with the greatest similarity node pair. The first merge step for the synthetic network is depicted in Figure 2. Further merge operations between supernodes simply merge the pair of distinct supernodes at a given step which contain the pair of ordinary nodes with the greatest GMT similarity (Figure 3). Note that the most GMT similar node pair might not be connected by an edge of the network, so that the method does not reduce to edge-collapsing with respect to an edge weighting on the original network. The precise steps are described in Algorithm 3. For notations used in Algorithms 1, 2 and 3, refer to Table 1.

2.3 Implementation and runtime complexity

A naive version of Algorithm 1 would iterate over all edge pairs, with complexity class $O(E^2)$ where E is the number of edges of the network. However, since only adjacent edge pairs are used, we instead iterate over the nodes and then over the pairs of its neighbors, with complexity class $O(ND^2)$ where N is the number of nodes and D is the

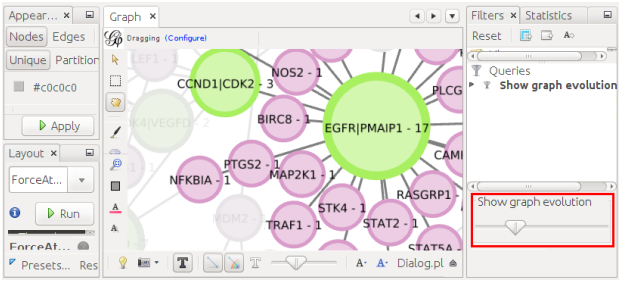


Figure 4: The Gephi plugin interface for exploring a supernode hierarchy. The dataset is a gene regulatory network. In this example, element size and color indicate the number of member nodes, and the labels indicate the most recently acquired members in a given supernode. They could also be used to indicate key node group statistics like the average Pearson correlation between the corresponding random variables, the average mutual information, or Bayes Information Criterion (BIC) values for prediction modeling of a given outcome/dependent variable of interest. The slider in the red bounding box is used to select different supernode hierarchical levels.

maximum degree over all nodes.

Normally, the computation of hierarchies requires that levels are computed successively, with each level depending on the previous one, precluding straightforward parallelization. In our case, finding the hierarchy itself is not computationally intensive, with the mixture modeling and GMT distances comprising the bulk of the computation. Both of these steps are classically parallelizable, the mixture modeling because it only depends on the initial node variable pair distributions, and the GMT distances because they only depend on the resulting list of mixture models. This makes our algorithm feasible for immediate computation. The discrete Earth Mover's Distance is performed with the R package 'emdist' [11]. The mixture modeling itself is performed with the R package 'mclust' [12]. In practice, the number P of mixture model populations has little effect on the overall output and performance as long as P lies in the approximate range from 3 to 10. If the number of node weighting samples is as low as a few hundred, it is not meaningful to choose P much greater than 10 anyway, since the number of data points per population should not be too low. High accuracy of the mixture model as a representation of the joint distribution of two given node variables is not essential for the purpose of inferring distances between the distributions from distances between the models.

Once the supernode hierarchy is computed, it is formatted for viewing in the Gephi open-source graph visualization software using a custom Java plugin. Figure 4 shows the typical appearance of the graphical user interface. The specific supernode hierarchical level for a given network is selected with a simple slider as shown in Figure 4. The labels are gene names of nodes most recently added to a given supernode, shown with a hyphen followed by the number of nodes contained in that supernode.

3. BACKGROUND ON OMT FOR GAUSSIAN MIXTURE MODELS

Table 1: Notation for Algorithms 1, 2, and 3

N	number of nodes
E	number of edges
S	number of node weightings
s	source function $[1, E] \to [1, N]$
t	target function $[1, E] \to [1, N]$
w	node weight matrix $[1, N] \times [1, S] \to \mathbb{R}$
P	Gaussian mixture modeling population number
GMM	Gaussian mixture model, 'mclust' R package [12]
$j \sim i$	nodes i and j are neighbors

Algorithm 1 GMT distance-based node similarities

```
function GMT SIMILARITIES(s,t,w,P)
    for e in [1,E] do
        Model(e) =GMM(w(s(e),−), w(t(e),−), P)
    end for
    list = {}
    for i in [1,N] do
        for j ∼ i do
            for k ∼ i, k > j do
                M1 = Model(e(j,i))
                M2 = Model(e(k,i))
                d =GMT Distance(M1,M2)
                list ← ((j,k),d)
            end for
        end for
    end for
    groups = group(list) by (j,k) value
    similarities = {}
    for group (j,k) in groups do
        similarities(j,k) =average d over group
    end for
    return similarities
end function
```

Algorithm 2 GMT distance functions

```
function GMT DISTANCE(M1,M2)
    D1 = probabilities(M1)                ▷ Vector size P
    D2 = probabilities(M2)                ▷ Vector size P
    for a in [1,P] do
        for b in [1,P] do
            G1 = Gaussian model(M1, a)
            G2 = Gaussian model(M2, b)
            cost(a,b) = OMT distance(M1(a),M2(b))
        end for
    end for
    return Earth Mover's Distance(D1, D2, cost)
end function
function OMT DISTANCE(G1,G2)
    v1 = mean(G1)
    v2 = mean(G2)
    Σ1 = covariance matrix(G1)
    Σ2 = covariance matrix(G2)
    D = sum of squared entries of v1 − v2
    D' = trace(Σ1 + Σ2 − 2(Σ1^(1/2) Σ2 Σ1^(1/2))^(1/2))
    return (D + D')^(1/2)
end function
```

A Gaussian mixture model is an important instance of

Algorithm 3 GMT supernode hierarchy

```
function GMT SUPERNODES(s,t,w,P)
    similarities = GMT Similarities(s, t, w, P)
    sort similarities descending
    supernodes = [1, N]                          ▷ Initialization
    for i in [1, N] do
        supernode containing(i) = i
    end for
    for (j, k) in domain(similarities) do
        s1 = supernode containing j
        s2 = supernode containing k
        s = UNION(s1, s2)
        supernode containing(j) = s
        supernode containing(k) = s
        supernodes ← s
    end for
    return supernodes
end function
```

the general mixture model structure, a structure that is commonly utilized to study properties of populations with several subgroups [5]. Formally, a *Gaussian mixture model* (GMM) is a probability density consisting of weighted linear combination of several Gaussian components, namely

$$\mu = q^1\pi^1 + q^2\pi^2 + \cdots + q^P\pi^P,$$

where each π^k is a Gaussian distribution and $q = (q^1, q^2, \ldots, q^P)^T$ is a probability vector. Here the finite number P stands for the number of components of μ.

Let μ_0, μ_1 be two Gaussian mixture models of the form

$$\mu_i = q_i^1\pi_i^1 + q_i^2\pi_i^2 + \cdots + q_i^{P_i}\pi_i^{P_i}, \quad i = 0, 1.$$

The distribution μ_i is equivalent to a discrete measure q_i with supports $\pi_i^1, \pi_i^2, \ldots, \pi_i^{P_i}$ for each $i = 0, 1$. The framework from [6] is based on the discrete OMT problem

$$\min_{\pi \in \Pi(q_0, q_1)} \sum_{i,j} c(i,j)\pi(i,j) \tag{1}$$

for these two discrete measures, where $\Pi(q_0, q_1)$ denotes the space of joint distributions with marginal distributions q_0 and q_1. The cost $c(i,j)$ is taken to be the 2-Wasserstein metric:

$$c(i,j) = W_2(\pi_0^i, \pi_1^j)$$

There is a closed formula for this metric:

$$\begin{aligned} W_2(\pi, \widetilde{\pi})^2 = \|m - \widetilde{m}\|^2 \\ + \operatorname{trace}(\Sigma + \widetilde{\Sigma} - 2(\Sigma^{1/2}\widetilde{\Sigma}\Sigma^{1/2})^{1/2}) \end{aligned} \tag{2}$$

where π and $\widetilde{\pi}$ are Gaussian distributions.

The discrete OMT problem (1) always has at least one solution, and letting π^* be a minimizer, we define

$$\text{GMT Distance}(\mu_0, \mu_1) = \sqrt{\sum_{i,j} c(i,j)\pi^*(i,j)}. \tag{3}$$

This formula, from [6], is the key formula underlying our algorithm.

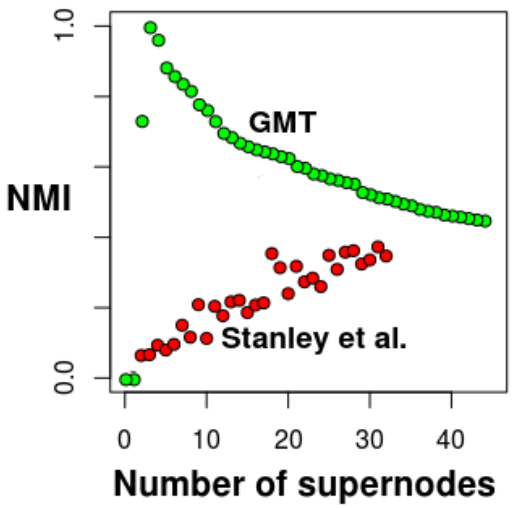


Figure 5: NMI values between the community structure provided by the GMT supernode hierarchy at a given level of complexity (indicated by the number of supernodes) and the $K = 3$ ground-truth communities of size 15 in the synthetic network shown in Figure 1. Also shown are the corresponding NMI values using the supernode-based compression algorithm of Stanley *et al.* [2] calculated on the same network. NMI measures the degree of concordance between two partitions, with NMI= 1 indicating perfect concordance.

4. EXPERIMENTS

All experiments were run on an Intel i3-3110M dual-core CPU processor with 4GB RAM. For the synthetic network (Figure 1) with K=3 communities, N=45 total nodes, and E=356 edges, the mixture modeling, GMT distance calculation, and hierarchy formation runtimes were respectively 9.5, 20.0, and 1.2 seconds. For the KEGG network (Figure 6) with N=187 nodes and E=275 edges, the runtimes were 3.60, 2.64, and 1.26 seconds.

4.1 Comparison with baseline method

Figure 5 shows Normalized Mutual Information (NMI) values comparing the GMT supernode hierarchy and the algorithm of Stanley *et al.* [2] on the synthetic dataset described in Figure 1. The NMI values for the Stanley algorithm are expected to be relatively low, since the supernode structures of that method are intended to be used as preprocessed input to more thorough community-detection algorithms like the Louvain algorithm or Stochastic Block Model fitting. The peak observed in the NMI plot for the GMT-based algorithm reflects the perfect concordance between the supernode hierarchy level with 3 supernodes and the known community structure with 3 communities. In general there are sharp a priori limits on the possible NMI values when the number of supernodes exceeds the number K of ground-truth communities.

4.2 Gene Regulatory Network Lung Cancer Dataset

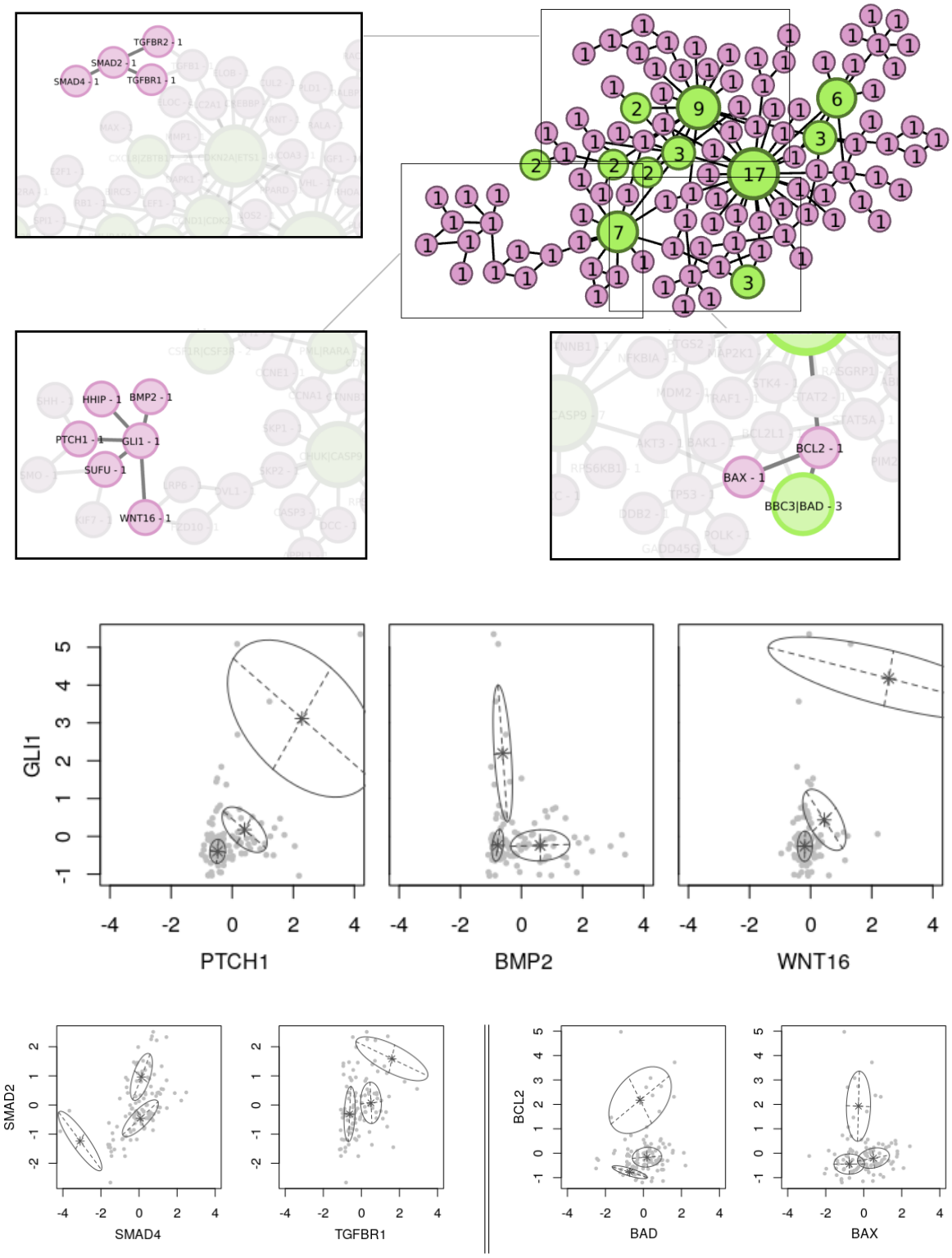


Figure 6: The KEGG cancer gene pathway network at a low or early level of the GMT supernode hierarchy. The 3 plots in the middle show joint distributions of z-scores of mRNA expression values for the GDSC lung cancer cell-line dataset for the "triad" motif ((PTCH1,WNT16),BMP2) with center node GLI1. The population centers and covariance matrix ellipses of the Gaussian mixture models with 3 populations are overlaid. See Table 2 for the pairwise GMT distances for each of the 3 pairs among the triad, and for the two pairs shown below. In the overview graphic, numbers indicate the number of nodes in a supernode.

Table 2: GMT distances for selected edge pairs which are adjacent along a center node in the gene regulatory network described in Figure 6. Listed with z-scores over all calculated GMT distances, for comparison. The first 3 rows and the next 3 rows show evidence of a "triad" motif. The last two rows are examples of low and high GMT distance, both illustrated in the plots in Figure 6.

Endpoint nodes		Center node	GMT distance	z-score
PTCH1	BMP2	GLI1	1.56	1.25
WNT16	BMP2	GLI1	2.07	2.69
WNT16	PTCH1	GLI1	0.65	-1.00
CCND1	RPS6KA5	MYC	1.92	2.30
CXCL8	RPS6KA5	MYC	1.75	1.84
CXCL8	CCND1	MYC	0.77	-0.70
SMAD4	TGFBR1	SMAD2	0.91	-0.31
BAD	BAX	MYC	1.64	1.57

The GMT supernode hierarchy was computed for the KEGG curated network of 187 cancer pathway genes with node weightings given by z-scores of mRNA expression values for the 108 lung cancer cell-lines provided by the GDSC project [9]. Table 2 shows selected exceptional values of the GMT distance among adjacent edge pairs for which the individual edges' node pairs have relatively high covariance. Unlike in the synthetic example, in the presence of noisy measurements it is relatively common for edges to have low covariance between their node variables.

The network in Figure 6 is a low level of the supernode hierarchy, constructed after only a few steps. A "triad" or "cycle" motif was found with central node GLI1 and three neighboring nodes PTCH1, WNT16, and BMP2. In this motif, (1) two of the edges both show significant correlation between their endpoint nodes, (2) the third edge shows significant correlation but of a different type, and (3) the three pairwise GMT distances reflect (1) and (2) with two large distances and one small distance. The joint distributions for the relevant node pairs are shown in Figure 6. An additional triad was found with central node MYC and three neighboring nodes CCND1, CXCL8, and RPS6KA5 (listed in Table 2 but not shown in the Figures).

Such motifs correspond to elementary operations of the hierarchy construction, specifying at the same time a collapse operation and two "do not collapse" operations for edge pairs, reminiscent of the "must link" and "cannot link" constraints of [3] but for edges rather than nodes. They should be expected in large quantities in cases where the node weightings evince non-trivial relationships between connected nodes.

The network in Figure 7 is a high level of the supernode hierarchy. In practice the typical situation for the highest levels is that one supernode comes to dominate the network, absorbing all others. So the other supernodes or ordinary nodes that remain are notable in that the joint distributions corresponding to the edges joining them to the large supernode S via some internal node N in S must be significantly different from the edges interior to the large supernode neighboring N. In this case one of the remaining supernodes is (PTCH1,WNT16), which is consistent with the previous observation that this pair belongs to a triad with a "do not collapse" tendency between the edges with endpoints PTCH1, WNT16 and a third edge.

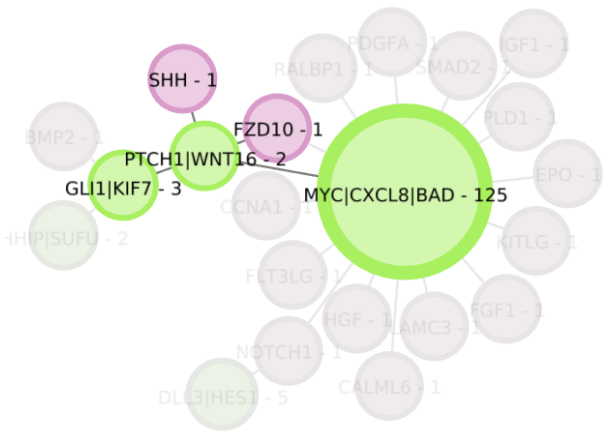


Figure 7: One of the last or highest levels of the GMT supernode hierarchy of the KEGG cancer pathway network using the GDSC lung cancer cell-line expression data.

5. CONCLUSION AND FUTURE WORK

We presented a supernode hierarchy construction based on the modified optimal mass transport methodology developed for Gaussian mixture models [6]. By representing a network at several levels of complexity with respect to functional relationships between edge pairs, it was shown to be an effective method for identifying structural features in real-world datasets. The one-shot parallel computation of the GMT metric allows real-time interactivity with the visualization tool.

There is much room for expansion of our basic framework. A finer theoretical analysis of the edge-local calculations may suggest a more informed version of the method which takes into account the entropy and mutual information of the random variables associated with the nodes. More practically, a pipeline supervised by prior knowledge of regions of interest in the network would dramatically increase the usefulness of the method as an exploratory tool. Further incorporation of Bayes Information Criterion (BIC) values for linear models fitted to response variables in terms of input variables delineated by each supernode could highlight outcome-relevant features in real-time. Several improvements to the GUI are possible, including graph layout defined by the hierarchy, automatic overlay of original data plots as in Figure 6, and the display of key representative elements of large supernodes selected by means of importance criteria defined by standard network measures. Though our algorithm is highly parallelizable, memory usage in the current R implementation has not been optimized and future versions will scale to much larger networks.

6. REFERENCES

[1] R. Albert and A. Barabási, "Statistical mechanics of complex networks," *Reviews of Modern Physics*, vol. 74, p. 1, 2002.

[2] N. Stanley, R. Kwitt, M. Niethammer, and P. J. Mucha, "Compressing networks with super nodes," *Scientific*

reports, vol. 8, no. 1, p. 10892, 2018.

[3] L. Yang, D. Jin, D. He, H. Fu, X. Cao, and F. Fogelman-Soulie, "Improving the efficiency and effectiveness of community detection via prior-induced equivalent super-network," *Scientific Reports*, vol. 7, no. 1, p. 634, 2017.

[4] M. Besta and T. Hoefler, "Survey and taxonomy of lossless graph compression and space-efficient graph representations," *arXiv preprint arXiv:1806.01799*, 2018.

[5] G. McLachlan and D. Peel, *Finite mixture models.* John Wiley & Sons, 2004.

[6] Y. Chen, T. T. Georgiou, and A. Tannenbaum, "Optimal transport for gaussian mixture models," *IEEE Access*, vol. 7, pp. 6269–6278, 2019.

[7] C. Villani, *Topics in Optimal Transportation.* American Mathematical Soc., 2003, no. 58.

[8] ——, *Optimal Transport: Old and New.* Springer, 2008, vol. 338.

[9] W. Yang, J. Soares, P. Greninger, E. J. Edelman, H. Lightfoot, S. Forbes, N. Bindal, D. Beare, J. A. Smith, I. R. Thompson, S. Ramaswamy, P. A. Futreal, D. A. Haber, M. R. Stratton, C. Benes, U. McDermott, and M. J. Garnett, "Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells," *Nucleic Acids Res.*, vol. 41, no. Database issue, pp. D955–961, Jan 2013.

[10] T. Georgiou, O. Michailovich, Y. Rathi, J. Malcolm, and T. A., "Distribution metrics and image segmentation," *Linear Algebra and Its Applications*, vol. 425, pp. 663–672, 2007.

[11] S. Urbanek and Y. Rubner, "emdist: Earth mover's distance," https://cran.r-project.org/web/packages/emdist/, 2012.

[12] L. Scrucca, M. Fop, T. B. Murphy, and A. E. Raftery, "mclust 5: clustering, classification and density estimation using Gaussian finite mixture models," *The R Journal*, vol. 8, no. 1, pp. 205–233, 2016. [Online]. Available: https://journal.r-project.org/archive/2016-1/scrucca-fop-murphy-etal.pdf